PRE-TEST ESTIMATORS OF THE INTERCEPT FOR A REGRESSION MODEL WITH MULTIVARIATE STUDENT-t ERRORS

Shahjahan Khan
Department of Mathematics & Computing
University of Southern Queensland
Toowoomba, Australia

and

A.K.Md. Ehsanes Saleh
Department of Mathematics & Statistics
Carleton University
Ottawa, Canada

ABSTRACT

In presence of an *uncertain prior information* about the slope parameter, the estimation of the intercept of a simple regression model with a multivariate Student-t error distribution is investigated. The *unrestricted*, *restricted* and *preliminary test* maximum likelihood estimators are defined. The expressions for the bias and the mean square error of the three estimators are provided and the *relative efficiencies* are analysed. A *maximin* criterion is established, and graphs and tables are constructed for different number of degrees of freedom (D.F.) as well as sample sizes. These tables of relative efficiencies can be used to determine a proper choice of the significance level of the preliminary test which in turn determines the choice of the estimator.

Key Words and Phrases: *Regression model, uncertain prior information, preliminary test estimator, multivariate t-distribution, inverted Gamma and non-central F distributions, incomplete Beta distribution, unrestricted and restricted maximum likelihood estimators, mean square errors, relative efficiency, maximin rule.*

AMS 1991 Subject Classification: Primary 62F30 and Secondary 62J05.

1. INTRODUCTION

Often in linear regression analysis, the distribution of the errors is assumed to be normal, and independent. In this paper a broader assumption is made, namely the errors are assumed to follow a family of Student-t distributions. Thus the joint distribution of the error components associated with the responses is a multivariate Student-t. The assumption of a t-model violates two basic assumptions of the normal model, namely, the normal distribution and the independence of the sample responses as under the t-model the sample responses are uncorrelated but not independent.

The problem at hand falls in the realm of the statistical inference problem with *uncertain prior information*. In the process of solving this kind of problems the "uncertain prior information" that appears in the form of a constraint is treated as the "nuisance parameter" and removed by using "Fisher's Recipe" of "testing it out". Such a problem was first addressed by Bancroft (1944) and the resulting estimator has been known in the literature as the *preliminary test* (PT) estimator. Since then the PT estimator has been studied by a host of authors, notably, Mosteller (1948), Kitagawa (1963), Han and Bancroft (1968), Saleh and Sen (1978), Saleh and Sen (1984), Saleh and Han (1990), and more recently, Saleh and Kibria (1993) and Wan (1994) to mention a few, under the normal theory.

The increasing criticism of the normal theory with its often unrealistic assumptions of independence and identicity as well as being non-robust has led researchers to find a better alternative among the class of symmetrical distributions. In many cases, the theoretical advantages and mathematical conveniences are negligible compared to the price paid in terms of loss of efficiency and precision under the normal theory. The concern was voiced by Fisher (1956, p.133) in the following words,"It is a noteworthy peculiarity of inductive inference that comparatively slight differences in the mathematical specification of a problem may have logically important effects on the inferences possible." Not surprisingly Fisher (1960, p.37) analysed Darwin's data under normal theory and later (p.46) assuming a symmetrical distribution. Fraser and Fick (1975) analysed the same data using a family of Student-t distributions. Obviously, the family of t-distributions represents a spectrum of symmetric densities ranging from the normal as the degrees of freedom parameter, $\nu_0 \to \infty$, down to the Cauchy when $\nu_0 = 1$, and to even thicker tailed sub-Cauchy distributions for $0 < \nu_0 < 1$.

Zellner (1976) revealed the fact that dependent but uncorrelated responses can be analysed by a Student-t model. He discussed the differences as well as the similarities of the results in both classical and Bayesian context for the normal and Student-t based models.

Fraser (1979, p.37) emphasised that the normal distribution is extremely short tailed and thus unrealistic as a sole distribution for variation. He also demonstrated the robustness of the Student-t family as opposed to the normal distribution based on empirical studies and analyses. His findings suggest that a Student-t model based analysis works reasonably well both for the normal and the Student-t model responses, but the same does not hold for the commonly normal based analysis (cf. Fraser, 1979, p.41).

In justifying the appropriateness and essence of the Student-t model Prucha and Kalajian (1984) have pointed out that the normal model based analysis (i) is generally very sensitive to deviations from its assumption, (ii) places too much weight on 'outliers', (iii) fails to utilize sample information beyond the first two moments, and (iv) appeals to a central limit theorem at most approximately, not exactly, normal.

In this paper, we consider a linear regression model,

$$y_j = \theta + \beta x_j + e_j; \qquad j = 1, 2, \cdots, n \tag{1.1}$$

where y_j is the j^{th} value of the dependent variable corresponding to x_j, a given fixed value of the independent variable; e_j is the error component associated with the response y_j; and θ and β are the intercept and slope parameters respectively of the model. It is assumed that the vector of the errors $\boldsymbol{e} = (e_1, e_2, \cdots, e_n)'$ is distributed according to the multivariate Student-t law with $E(\boldsymbol{e}) = \boldsymbol{o}$ and $E(\boldsymbol{ee}') = {\sigma_e}^2 I_n$ where σ_e^1 is the common variance of e_j' s, $j = 1, 2, \cdots, n$ and I_n is the identity matrix of order n. The class of Student-t distributions with varying degrees of freedom can be expressed as a variance mixture of normal distributions as follows:

$$f(\boldsymbol{e}; \sigma^2, \nu_0) = \int_0^\infty f_N(\boldsymbol{e}) f(\tau) d\tau \tag{1.2}$$

where $f_N(\boldsymbol{e})$ is the p.d.f. of $\boldsymbol{e}$ when $\boldsymbol{e} \sim N(0,\ \tau^2 I_n)$; and $f(\tau)$ is the p.d.f. of an *Inverted Gamma* (IG) distribution with scale parameter σ^2 and degrees of freedom parameter ν_0, that is, $\tau \sim \text{IG}(\sigma^2, \nu_0)$. Therefore, we obtian the joint density of $(e_1, e_2, \cdots, e_n)'$ as

$$f(\boldsymbol{e}; \sigma, \nu_0) = \frac{h(\nu_0)}{\sigma^n} \left[\nu_0 + \frac{1}{\sigma^2} \sum_{j=1}^n e_j^2\right]^{-\frac{\nu_0+n}{2}} \tag{1.3}$$

with $h(\nu_0) = \frac{\nu_0^{\frac{\nu_0}{2}} \Gamma(\frac{\nu_0+n}{2})}{\pi^{\frac{n}{2}} \Gamma(\frac{\nu_0}{2})}$, as the normalizing constant, $\sigma_e^2 = \frac{\nu_0}{\nu_0 - 2} \sigma^2 > 0$, $\nu_0 \geq 3$, and $-\infty < \theta < \infty$.

It is well known that the above density function approaches to the normal form as $\nu_0 \to \infty$, and when $\nu_0 = 1$, it becomes Cauchy. Furthermore, the marginal distribution of each component of $\boldsymbol{e}$ is univariate Student-t. Also, for smaller values of ν_0, this distribution has thicker tails than that under normal distribution.

Our choice of the Student-t family thus includes a class of symmetrical distributions with variable tail thickness. Unlike the normal distribution, it is capable of handling 'outliers' as well as dependent but uncorrelated responses. Further support for the application of the Student-t model may be found in Haq and Khan (1990), and Khan and Haq (1994).

In this paper, we consider the problem of estimating the intercept parameter, θ based on the observed sample observations $(y_1, y_2, \cdots, y_n)$ as specified in the model (1.1) when an *uncertain prior information* in the form of a null hypothesis $H_0 : \beta = \beta_0$ is available. We define three different estimators, namely, *unrestricted*, *shrinkage restricted*, and *shrinkage preliminary test* estimators of θ and study their properties based on both the unbiasedness and mean square error (m.s.e.) criterion. In the presence of an uncertain prior information the usual procedure is to pre-test H_0 before the actual estimation of the parameter.

The problem of estimation as well as the sampling properties of the estimators for the linear regression model when an uncertain prior information exists have been widely investigated in the literature [see for instance, Judge and Bock (1978), Griffiths et al. (1992) and the references there in] under the normal theory. As discussed above, there is an increasing evidence in the literature that in many cases the set of data may have been generated by processes whose distributions have higher *kurtosis*, that is, heavier tails than the normal distributions. Giles (1991) cited a number of references from the commodity and financial markets study where the underlying distributions are symmetrical but not normal. Several of those authors revealed the fact that in many real life data the Student-t distribution model fits far better than the normal based model. For example, see Praetz (1972), and Blattberg and Gonedes (1974). The implications of these applied studies resulted in the use of the Student-t distribution of the error terms in the regression model by many authors. Recently, statistical analyses of the linear regression model based on the Student-t distribution have been pursued by Giles (1991), Khan and Haq (1994), and Khan and Saleh (1995) to mention a few.

In the next section we define three different estimators of θ. Some useful theorems for the computation of the m.s.e. of the estimators are given in section 3. The expressions for the bias and the m.s.e. are obtained in section 4, while the relative efficiency for the estimators are derived in section 5. Comparisons among the estimators and recommendations are made in section 6. Some concluding remarks are included in section 7.

2. Estimators of the Intercept

In this section, our objective is to define various estimators of θ based on the sample observations $(y_1, y_2, \cdots, y_n)$ having the joint density as given in (1.3) when it is suspected that the null hypothesis $H_0 : \beta = \beta_0$ may hold, but not sure.

It is well known (cf. Zellner, 1976) that the maximum likelihood estimators (m.l.e.) of β, θ and σ^2 for the simple regression model with errors having multivariate Student-t distribution are

$$\begin{aligned}\tilde{\beta}_n =& \frac{\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{n}(x_j - \bar{x})^2},\\ \tilde{\theta}_n =& \bar{y} - \tilde{\beta}\bar{x}, \quad \text{and} \\ \tilde{\sigma}^2 =& s^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \tilde{\theta} - \tilde{\beta}x_j)^2\end{aligned} \tag{2.1}$$

respectively, where $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ and $\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$. The estimator $\tilde{\theta}_n$, as defined above, is a function of the maximum likelihood estimator of β and will be called the *unrestricted* estimator of θ. If the null hypothesis is true, then the natural estimator of β is β_0. However, since we are not sure about the tenability of the null hypothesis, we make use of the available sample information in the estimation of θ. Thus taking a convex combination of β_0 and $\tilde{\beta}_n$ we define the following *shrinkage restricted* (sr) estimator of β as

$$\hat{\beta}_n^{sr} = c\beta_0 + (1-c)\tilde{\beta}_n, \qquad 0 \le c \le 1 \tag{2.2}$$

where we call c as the *degree of confidence* in the null hypothesis. The estimator $\hat{\beta}_n^{sr}$ is known as the *shrinkage restricted* estimator of β. Thus for every possible value of c the estimator in (2.2) will produce a different estimator. Obviously, our estimator of β would be $\frac{1}{2}(\beta_0 + \tilde{\beta}_n)$ if $c = \frac{1}{2}$. Clearly, the *shrinkage restricted* estimator in (2.2) is a compromise between the two extreme estimators, one based on the null hypothesis, totally ignoring the sample information, and the other based on the sample observations alone, disregarding the *uncertain prior information* in the form of the null hypothesis. Thus we define *shrinkage restricted* estimator of θ as

$$\hat{\theta}_n^{sr} = \bar{y} - [c\beta_0 + (1-c)\tilde{\beta}_n]\bar{x}. \tag{2.3}$$

In applications, usually the null hypothesis is suspicious. In such a case we follow the "Fisher's Recipe" to remove the suspicion by testing H_0 out through an appropriate test statistic. Using Zellner's (1976) idea we consider the test statistic

$$F = \frac{ns_x^2(\tilde{\beta}_n - \beta_0)^2}{s_e^2} \tag{2.4}$$

to test $H_0 : \beta = \beta_0$, where

$$\begin{aligned} ns_x^2 &= \sum_{j=1}^{n} (x_j - \bar{x})^2 \quad \text{and} \\ (n-2)s_e^2 &= \sum_{j=1}^{n} \Big\{ (y_j - \bar{y}) - \tilde{\beta}_n (x_j - \bar{x}) \Big\}^2 . \end{aligned} \tag{2.5}$$

Under H_0, F in (2.4) follows a central F-distribution with $(1,\ m)$ D.F. where $m = n - 2$.

Thus, for the estimation of θ when $H_0 : \beta = \beta_0$ is suspicious, we carry out an F-test and choose $\tilde{\theta}_n$ if H_0 is not tenable and $\hat{\theta}_n^{sr}$ if H_0 is accepted at a prescribed level of significance, say, α $(0 < \alpha < 1)$. Then, we have the *shrinkage preliminary test* (spt) estimator of θ as follows:

$$\begin{aligned} \hat{\theta}_n^{spt} &= \hat{\theta}_n^{sr} I(F < F_\alpha) + \tilde{\theta}_n I(F \geq F_\alpha) \\ &= \tilde{\theta}_n - (\tilde{\theta}_n - \hat{\theta}_n^{sr}) I(F < F_\alpha) \\ &= \tilde{\theta}_n + c(\tilde{\beta}_n - \beta_0)\bar{x} I(F < F_\alpha) \end{aligned} \tag{2.6}$$

where $I(\cdot)$ is the indicator function which assumes values either 0 or 1, F_α is the $(1-\alpha)^{th}$ quantile of a central F-distribution with $(1,\ m)$ D.F. If $c = 1$, we obtain the ordinary *preliminary test* (pt) estimator:

$$\hat{\theta}_n^{pt} = \tilde{\theta}_n + (\tilde{\beta}_n - \beta_0)\bar{x} I(F < F_\alpha). \tag{2.7}$$

In the foregoing section, we have defined three estimators of θ, namely, $\tilde{\theta}_n$, the *unrestricted* m.l.e.; $\hat{\theta}_n^{sr}$, the *shrinkage restricted* m.l.e.; and $\hat{\theta}_n^{spt}$, the *shrinkage preliminary test* m.l.e.

To determine the power function of the test statistic under the alternative hypothesis, $H_1 : \beta \neq \beta_0$ we have some preliminaries in section 3.

3. Some Preliminaries

Theorem 3.1. Suppose $Y_1,\ Y_2, \cdots, Y_n$ are identically and independently distributed as $N(\theta^*,\ \tau^2)$ with $\theta^* = \mathrm{E}(Y_j)$, and τ follows an Inverted Gamma (IG) distribution with parameters $(\nu_0,\ \sigma^2)$ given by

$$f(\tau|\nu_0, \sigma^2) = \left\{ \frac{2}{\Gamma(\frac{\nu_0}{2})} \right\} \left\{ \frac{\nu_0 \sigma^2}{2} \right\}^{\frac{\nu_0}{2}} \tau^{-(\nu_0+1)} e^{-\frac{\nu_0\sigma^2}{2\tau^2}} . \tag{3.1}$$

Then the joint distribution of $Y_1,\ Y_2, \cdots, Y_n$ is given by

$$f(\boldsymbol{y}; \theta^*, \sigma, \nu_0) = \frac{h(\nu_0)}{\sigma^n} \Big[\nu_0 + \frac{1}{\sigma^2} \sum_{j=1}^{n} (y_j - \theta^*)^2 \Big]^{-\frac{\nu+n}{2}} . \tag{3.2}$$

Proof: Completing the following integration

$$\int_0^\infty (2\pi\tau^2)^{-\frac{n}{2}} e^{-\frac{1}{2\tau^2}\sum_{j=1}^n (y_j-\theta^*)^2} f(\tau|\nu_0,\sigma)d\tau \tag{3.3}$$

we get (3.2).

Theorem 3.2. The distribution of $g = m \times F$, where F is given by (2.4) under $H_1 : \beta \neq \beta_0$ follows the distribution defined by the following density

$$\begin{aligned} f(g) = \frac{2}{\nu_0 - 2}\sum_{r\geq 0}\Bigg\{ & \frac{\Gamma(r+\frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2}-1)} \frac{(\frac{\Delta^*}{\nu_0-2})^r}{\Big(1+\frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0+2r)}} \\ & \times \frac{\Gamma(r+\frac{n}{2})}{\Gamma(r+\frac{1}{2})\Gamma(\frac{n-1}{2})} \frac{g^{r-\frac{1}{2}}}{\Big(1+g\Big)^{r+\frac{n}{2}}}\Bigg\} \end{aligned} \tag{3.4}$$

where $\Delta^* = \frac{\delta^2}{\sigma_*^2}$ in which $\delta^2 = (\beta-\beta_0)^2$ and $\sigma_*^2 = \frac{\nu_0}{\nu_0-2}\sigma^2$.

Theorem 3.3. The distribution function (d.f.) of g is given by

$$\begin{aligned} G_{1,m}^{(1)}(F_0;\Delta^*) = \frac{2}{\nu_0 - 2}\sum_{r\geq 0}\Bigg\{ & \frac{\Gamma(r+\frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2}-1)} \frac{(\frac{\Delta^*}{\nu_0-2})^r}{\Big(1+\frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0+2r)}} \\ & \times I_u\Big(r+\frac{1}{2};\ \frac{m}{2}\Big)\Bigg\} \end{aligned} \tag{3.5}$$

where $I_u\Big(r+\frac{1}{2};\ \frac{m}{2}\Big)$ is the *incomplete Beta function* with $u = \frac{F_{1,m}(\alpha)}{m+F_{1,m}(\alpha)}$ and F_0 is the value of the $F_{1,m}$ variable such that $(1-\alpha)\times 100$ percent area under the central $F_{1,m}$ distribution curve is to the left of F_0 for given values of α and m. The proof follows from the straight forward expectation of the non-central F-distribution with respect to the *Inverted Gamma* distribution with parameters $(\nu_0,\ \sigma^2)$.

Now observe that

$$\lim_{\nu_0\to\infty} P(F \leq F_0) = G_{1,m}(F_0;\Delta) \tag{3.6}$$

which is the distribution function (d.f.) of a non-central F-distribution with $(1,\ m)$ D.F. and non-centrality parameter $\Delta = \frac{\delta^2}{\sigma^2}$ since

$$\begin{aligned} &\lim_{\nu_0\to\infty}\Bigg\{\frac{\Gamma(r+\frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2}-1)} \frac{(\frac{\Delta^*}{\nu_0-2})^r}{\Big(1+\frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0+2r)}}\Bigg\} = \frac{(\frac{\Delta}{2})^r e^{-\frac{\Delta}{2}}}{\Gamma(r+1)} \\ &\text{and} \quad \lim_{\nu_0\to\infty}\sigma_*^2 = \sigma^2. \end{aligned} \tag{3.7}$$

Therefore, the results obtained in this paper will remain valid for the normal based model as a special case when $\nu_0 \to \infty$.

In the next section, the expressions for the bias and the mean square error of the three estimators, defined in the previous section, are provided.

4. Bias and M.S.E. of Estimators

The bias and the m.s.e. of the *unrestricted* estimator of θ are

$$\begin{aligned} B_1(\tilde{\theta}_n) &= E\{\tilde{\theta}_n - \theta\} = 0, \text{ and} \\ M_1(\tilde{\theta}_n) &= nE\{\tilde{\theta}_n - \theta\}^2 = \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big(1 + \frac{\bar{x}^2}{s_x^2}\Big) \end{aligned} \tag{4.1}$$

respectively. Similarly, for the *shrinkage restricted* estimator of θ the expressions for the bias and the m.s.e. are

$$\begin{aligned} B_2(\hat{\theta}_n^{sr}) &= \sqrt{n}E\{\hat{\theta}_n^{sr} - \theta\} = -c\delta, \text{ and} \\ M_2(\hat{\theta}_n^{sr}) &= nE\{\hat{\theta}_n^{sr} - \theta\}^2 = \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big[1 + (1-c)^2\Big(\frac{\bar{x}^2}{s_x^2}\Big) + c^2\bar{x}^2\Delta^*\Big] \end{aligned} \tag{4.2}$$

respectively, where $\Delta^* = \frac{\delta^2}{\frac{\nu_0}{\nu_0-2}\sigma^2}$ in which $\delta = \sqrt{n}(\beta - \beta_0)$.

Finally, the expressions of the bias and the m.s.e. of the *shrinkage preliminary test* (spt) estimator of θ are

$$\begin{aligned} B_3(\hat{\theta}_n^{spt}) &= \sqrt{n}E\{\hat{\theta}_n^{spt} - \theta\} = -c\delta G_{3,m}^{(1)}(l_\alpha; \Delta^*), \text{ and} \\ M_3(\hat{\theta}_n^{spt}) &= nE\{\hat{\theta}_n^{spt} - \theta\}^2 \\ &= \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big[\Big(1 + \frac{\bar{x}^2}{s_x^2}\Big) - c(2-c)\Big(\frac{\bar{x}^2}{s_x^2}\Big)G_{3,m}^{(2)}(l_\alpha; \Delta^*) \\ &\qquad + c\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Delta^*\Big\{2G_{3,m}^{(1)}(l_\alpha; \Delta^*) - (2-c)G_{5,m}^{(1)}(l_\alpha^*; \Delta^*)\Big\}\Big] \end{aligned} \tag{4.3}$$

respectively, where $l_\alpha = \frac{1}{3}F_{1,m}(\alpha)$, $l_\alpha^* = \frac{1}{5}F_{1,m}(\alpha)$, and

$$\begin{aligned} G_{3,m}^{(1)}(l_\alpha; \Delta^*) = \frac{2}{\nu_0 - 2}\sum_{r\geq 0}\Big\{&\frac{\Gamma(r + \frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2} - 1)}\frac{(\frac{\Delta^*}{\nu_0 - 2})^r}{\Big(1 + \frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0 + 2r)}} \\ &\times I_{u_1}\Big(r + \frac{3}{2}; \frac{m}{2}\Big)\Big\} \end{aligned} \tag{4.4}$$

in which $u_1 = \frac{3l_\alpha}{m + 3l_\alpha}$,

$$\begin{aligned} G_{5,m}^{(1)}(l_\alpha^*; \Delta^*) = \frac{2}{\nu_0 - 2}\sum_{r\geq 0}\Big\{&\frac{\Gamma(r + \frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2} - 1)}\frac{(\frac{\Delta^*}{\nu_0 - 2})^r}{\Big(1 + \frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0 + 2r)}} \\ &\times I_{u_2}\Big(r + \frac{5}{2}; \frac{m}{2}\Big)\Big\} \end{aligned} \tag{4.5}$$

in which $u_2 = \frac{5l_\alpha}{m + 5l_\alpha}$, and

$$\begin{aligned} G_{3,m}^{(2)}(l_\alpha; \Delta^*) = \frac{2}{\nu_0 - 4}\sum_{r\geq 0}\Big\{&\frac{\Gamma(r - 1 + \frac{\nu_0}{2})}{\Gamma(r+1)\Gamma(\frac{\nu_0}{2} - 2)}\frac{(\frac{\Delta^*}{\nu_0 - 2})^r}{\Big(1 + \frac{\Delta^*}{\nu_0-2}\Big)^{\frac{1}{2}(\nu_0 + 2r - 2)}} \\ &\times I_{u_1}\Big(r + \frac{3}{2}; \frac{m}{2}\Big)\Big\} \end{aligned} \tag{4.6}$$

in which $u_1 = \frac{3l_\alpha}{m+3l_\alpha}$.

With the results obtained in this section, we provide the relative efficiencies of the estimators in the next section.

5. Relative Efficiency Expressions

The relative efficiency (R.E.) of $\hat{\theta}_n^{sr}$ with respect to $\tilde{\theta}_n$ is given by the following expression

$$E_1(c,\Delta^*) = \left(1+\frac{\bar{x}^2}{s_x^2}\right)\left[1+(1-c)^2\left(\frac{\bar{x}^2}{s_x^2}\right)+c^2\bar{x}^2\Delta^*\right]^{-1}, \qquad 0\le c\le 1, \tag{5.1}$$

and that of $\hat{\theta}_n^{spt}$ with respect to $\tilde{\theta}_n$ by

$$E_2(c,\alpha,\Delta^*) = \Big[1+\phi(\Delta^*)\Big]^{-1} \tag{5.2}$$

where

$$\begin{aligned}\phi(\Delta^*) =&\left(1+\frac{\bar{x}^2}{s_x^2}\right)^{-1}\Bigg[c\left(\frac{\bar{x}^2}{s_x^2}\right)\Delta^*\Big\{2G_{3,m}^{(1)}(l_\alpha;\Delta^*)\\ &-(2-c)\left(\frac{\bar{x}^2}{s_x^2}\right)G_{5,m}^{(1)}(l_\alpha^*;\Delta^*)\Big\}-c(2-c)G_{3,m}^{(2)}(l_\alpha;\Delta^*)\Bigg]^{-1}.\end{aligned} \tag{5.3}$$

Finally, the expression for the relative efficiency (R.E.) of $\hat{\theta}_n^{spt}$ with respect to $\hat{\theta}_n^{sr}$ is found to be

$$\begin{aligned}E_3(c,\alpha,\Delta^*) =&\left[1+(1-c)^2\left(\frac{\bar{x}^2}{s_x^2}\right)+c^2\bar{x}^2\Delta^*\right]\\ &\times\left(1+\frac{\bar{x}^2}{s_x^2}\right)^{-1}\Big[1+\phi(\Delta^*)\Big]^{-1}\end{aligned} \tag{5.4}$$

where $\phi(\Delta^*)$ is the same as defined in (5.3).

Based on the relative efficiency of the estimators, comparisons and recommendations are made in the following subsection.

6. Comparisons and Recommendations

First, we present the dominance picture of the three estimators under the null hypothesis in the following theorem.

Theorem 6.1. Under the $H_0 : \beta = \beta_0$ the dominance picture of the estimators is

$$\hat{\theta}_n^{sr} \prec \hat{\theta}_n^{spt} \prec \tilde{\theta}_n. \tag{6.1}$$

Proof: Consider the m.s.e. differences

(i)

$$M_1 - M_2 = \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Big[1-(1-c)^2\Big] > 0 \tag{6.2}$$

that is, $\hat{\theta}_n^{sr}$ is better than $\tilde{\theta}_n$ under H_0.

(ii)

$$M_1 - M_3 = \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Big[c(2-c)G_{3,m}^{(2)}(l_\alpha;0)\Big] > 0 \tag{6.3}$$

that is, $\hat{\theta}_n^{spt}$ is better than $\tilde{\theta}_n$ under H_0.

(iii)

$$M_3 - M_2 = \frac{\nu_0}{\nu_0 - 2}\sigma^2\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Big[c(2-c)\Big\{1-G_{3,m}^{(2)}(l_\alpha;0)\Big\}\Big] > 0 \tag{6.4}$$

that is, $\hat{\theta}_n^{sr}$ is better than $\hat{\theta}_n^{spt}$ under H_0.

Comparison Between $\hat{\theta}_n^{sr}$ and $\tilde{\theta}_n$

First, note that for a given set of x-values, both $\bar{x}$ and s_x^2 are known and fixed. Consider the relative efficiency expression $E_1(c,\Delta^*)$. For a fixed c, it is a decreasing function of Δ^*. At $\Delta^* = 0$,

$$E_1(c,0) = \Big(1+\frac{\bar{x}^2}{s_x^2}\Big)\Big[1+(1-c)^2\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Big]^{-1}, \qquad 0 \le c \le 1, \tag{6.5}$$

which becomes 1 when $c=0$ and $\Big(1+\frac{\bar{x}^2}{s_x^2}\Big)\Big[1+\frac{1}{2}\Big(\frac{\bar{x}^2}{s_x^2}\Big)\Big]^{-1}$ when $c=\frac{1}{2}$. As $\Delta^* \to \infty$, $E_1(c,\Delta^*) \to 0$ for $c \neq 0$. However, for $0 \le \Delta^* \le \frac{2-c}{c}$, $\hat{\theta}_n^{sr}$ performs better than $\tilde{\theta}_n$. For $c = \frac{1}{2}$, $\frac{2-c}{c} = 3$, thus for $0 \le \Delta^* \le 3$, $\hat{\theta}_n^{sr}$ performs better than $\tilde{\theta}_n$. If $c = 1$, the range of Δ^* is 0 to 1. Thus $\hat{\theta}_n^{sr}$ has a wider performance range than the ordinary PT estimator, $\hat{\theta}_n^{pt}$ for small c values. Outside this range $\tilde{\theta}_n$ performs better.

Comparison Between $\hat{\theta}_n^{spt}$ and $\tilde{\theta}_n$

Once again, consider the R.E. expression, $E_2(c,\alpha,\Delta^*)$ as given in (5.2). For a fixed c, it is a function of (α,Δ^*). The function has its maximum at $\Delta^* = 0$ with the value $E_2^*(c,\alpha,0) = E_2^*$ given by

$$E_2^* = \Big[1 - c(2-c)\Big(1+\frac{\bar{x}^2}{s_x^2}\Big)^{-1}G_{3,m}^{(2)}(l_\alpha;0)\Big]^{-1} > 1. \tag{6.6}$$

The function $E_2(c,\alpha,\Delta^*)$ decreases as Δ^* increases crossing the line $E_2(c,\alpha,\Delta^*) = 1$ to a maximum $E_2^0(c,\alpha,\Delta^*)$ at $\Delta^0 = \Delta^*$, then increases towards 1 as $\Delta^* \to \infty$. Now, for $\Delta^* = 0$ and α varying, we have

$$\max_{0\le\alpha\le1} E_2(c,\alpha,0) = E_2(c,0,0) = [1-c(2-c)]^{-1}. \tag{6.7}$$

The value of $E_2(c,\alpha,0)$ decreases as α increases. On the other hand, when $\alpha=0$ and Δ^* varies, then the curves $E_2(c,0,\Delta^*)$ and $E_2(c,1,\Delta^*)=1$ intersect at $\Delta^*=\frac{2-c}{c}$. For general α, the functions $E_2(c,\alpha_1,\Delta^*)$ and $E_2(c,\alpha_2,\Delta^*)$ will always intersect in the interval $0\le\Delta^*\le\frac{2-c}{c}$. The value of Δ^* at the intersection decreases as α increases. Therefore, for two different α, say α_1 and α_2, the functions $E_2(c,\alpha_1,\Delta^*)$ and $E_2(c,\alpha_2,\Delta^*)$ will never intersect above $E_2(c,1,\Delta^*)=1$. Some graphs are shown to illustrate the phenomenon.

In order to choose an estimator with maximum relative efficiency, we adopt the following rule:

If it is known that $\Delta^*\in[0,\frac{2-c}{c}]$, the estimator $\hat{\theta}_n^{sr}$ is chosen since $E_2(c,\alpha,\Delta^*)$ is largest in this interval. However, Δ^* is generally unknown and in which case there is no way of choosing an *uniformly best* estimator. Thus we pre-assign a prescribed R.E.-value, say, E^0 which is tolerable. Then consider the set $A=\{\alpha|\ E_2(c,\alpha,\Delta^*)\ge E^0\}$ and try to choose an estimator which maximises $E_2(c,\alpha,\Delta^*)$ over all $\alpha\in A$ and all Δ^*. That is, we solve for α such that

$$\max_{\alpha\in A}\ \min_{\Delta^*} E_2(c,\alpha,\Delta^*)=E^0. \tag{6.8}$$

Hence we have a *maximin* rule for the optimum choice of the level of significance for the *shrinkage preliminary test* estimator, $\hat{\theta}_n^{spt}$. Tables of R.E., both maximum (E^*) and minimum (E^0), and the value of Δ^* at which the minimum occurs (Δ^0) for a given α are reported at the end of this section.

Discuss an example from the table with $\nu_0=3,6,9,12,15$.

7. CONCLUDING REMARKS

In this paper, we have studied the properties of the *unrestricted, shrinkage restricted* and *preliminary test* estimators of the intercept parameter of the linear regression model with a class of Student-t errors. Since the Student-t model with ν_0 D.F. encompasses a class of symmetrical distributions which includes the normal as $\nu_0\to\infty$ as well as other wider tailed distributions so the proposed estimators in this paper are robust in this class of regression models. In this study, we find the behaviour of the estimators similar to those in the case of the normal model. The differences in the relative efficiencies of the estimators are shown by the graphs for varied values of ν_0. The decision rule for selecting an optimal α, the level of significance is also shown using tabular values of maximum (E^*) and minimum (E^0) relative efficiencies for a range of ν_0-values. It is expected that the results obtained in this paper will be preferable to normal theory results by the practitioners as it includes the normal based results as a special case and many others in the family of elliptically symmetric distributions. Moreover, the results obtained under the Student-t model are valid for the normal model but not vice-versa.

REFERENCES

Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Statist.*, **15**, 190-204.

Blattberg, R.C. and Gonedes, N.J. (1974). A comparison of the stable and Student distribution as statistical model for stock prices. *Journal of Business*, **47**, 244-280.

Fisher, R.A. (1956). Statistical Methods Scientific Inference. Oliver and Boyd, London.

Fisher, R.A. (1960). The Design of Experiments. 7^{th} ed. Hafner, New York.

Fraser, D.A.S. (1979). Inference and Linear Models. McGrow Hill, New York.

Fraser D.A.S. and Fick, G.H. (1975). Necessary analysis and its implementation. Proc. Symposium on Statistics and Related Topics, A.K.Md.E. Saleh (ed.), Carleton University, Canada, 5.01-5.30.

Giles, J.A. (1991). Pre-testing for linear restrictions in a regression model with spherically symmetric disturbances. *Journal of Econometrics*, **50**, 377-398.

Griffiths, W.E., Lutkepohl, H., and Bock, M.E. (1992). Readings in Econometric Theory and Practice: a volume in honor of George Judge. (ed.) North-Holland, New York.

Han, C.P. and Bancroft, T.A. (1968). On pooling means when variance is unknown. *Jou. Amer. Statist. Assoc.*, **63**, 1333-1342.

Haq, M.S. and Khan, S. (1990). Prediction distribution for linear model with Multivariate Student-t errors distribution. *Communications in Statistics, Theory & Methods,* **19**, 4705-4712.

Judge, G.G. and Bock, M.E. (1978). Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics. North-Holland, Amsterdam.

Khan, S. and Haq, M.S. (1994). Predictive inference for the multilinear model with errors having multivariate t-distribution and first order auto-correlation structure. *Sankhya, B*, **56**, 95-106.

Khan, S. and Saleh, A.K.Md.E.(1995). Preliminary test estimators of the mean for sampling from some Student-t populations. *Faculty of Sciences Wroking Paper Series, University of Southern Queensland, Australia.*

Kitagawa, T, (1948). Estimation after preliminary test of significance. *University of California Publications in Statistics*, Vol. B, 147-186.

Lobato, I. N. (2001). Testing That a Dependent Process Is Uncorrelated, Journal of the American Statistical Association, Volume 96(455), 1066-1076(11)

Mosteller, F. (1948). On pooling data. *Jou. Amer. Statist. Assoc.*, **43**, 231-242.

Praetz, P.D. (1972). The distribution of share price changes. *Journal of Business*, **45**, 49-55.

Prucha, I.R. and Kelejian, H.H. (1984). The structure of simultaneous equation estimators: A generalization towards non-normal disturbances. *C*, **52**, 721-736.

Saleh, A.K.Md.E. and Sen, P.K. (1978). Non-parametric estimation of location parameter after a preliminary test on regression. *Ann. Statist.*, **6**, 154-168.

Saleh, A.K.Md.E. and Kibria, B.M.G. (1993). Performance of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics, Theory & Methods,* **22**, 2747-2764.

Saleh, A.K.Md.E. and Han, C.P. (1990). Shrinkage estimation in regression analysis. *Estadistica,* **42**, 40-63.

Wan, A.T.K. (1994). The sampling preformance of inequality restricted and pre-test estimators in a mis-specified linear model. *Australian Journal of Statistics,* **36**(3), 313-325.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error term. *Jou. Amer. Statist. Assoc.* **60**, 608-616.